© Grace Scientific Publishing

Journal of
Statistical Theory and Practice
Volume 2, No.3, September 2008

On Optimal Uniform Deconvolution

Andrey Feuerverger, *Department of Statistics, University of Toronto, Toronto, Ontario M5S 3G3, Canada. Email: andrey@utstat.toronto.edu*

Peter T. Kim, *Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario N1G 2W1, Canada. Email: pkim@uoguelph.ca*

Jiayang Sun, *Department of Statistics, Case Western Reserve University, Cleveland, Ohio 44106, U.S.A. Email: jiayang@sun.cwru.edu*

Received: July 26, 2007 *Revised:* April 2, 2008

Abstract

This paper concerns the nonstandard problem of uniform deconvolution for nonperiodic functions over the real line. New algorithms are developed for this nonstandard statistical problem and integrated mean squared error bounds are established. We show that the upper bound of the integrated mean squared error for our new procedure is the same as for the standard case; hence these new estimators attain the lower bound minimax, and hence optimal, rate of convergence. Our method has potential applications to such problems as the deblurring of optical images which have been subjected to uniform motion over a finite interval of time. We also treat the case when the support of the uniform is not given and must be estimated. The numerical properties of our algorithms are demonstrated and shown to be well behaved.

AMS Subject Classification: Primary 62G20. Secondary 65R32.

Key-words: Deconvolution; characteristic functions; Fourier inversion; rate optimality; uniform distribution.

1. Introduction and summary

When signals pass through filters or systems that are linear and time invariant, the resulting output is a convolution of the original signal with the filter's impulse response function. In typical applications, the output signals are observable, but the input signals are not. For this and related reasons, deconvolution is an important problem in signal and image processing, as well as in many engineering applications. This problem also falls under the broader category of statistical inverse problems.

* 1559-8608/08-3/$5 + $1pp – see inside front cover

The literature related to deconvolution – and associated statistical inverse problems, such as when noise is subsequently added to the system – is therefore extensive. Within statistics, convolution of densities occurs, for example, in measurement error and related statistical models, in which independent random variables are added together. The prototypical problem involves contaminated independent observations composed of the true observations plus independent additive noise. In this setting, which is the one that motivates this work, the objective is to recover the density of the true observations from the contaminated observations, given the noise distribution or some defining characteristics thereof.

Statistical deconvolution problems have been studied by many authors, both in terms of the framework above, and in the broader scope of statistical inverse problems. In so doing a standard framework has emerged with some of the deconvolution papers being Carroll and Hall [4], Fan [9], Stefanski and Carroll [18], Zhang [20], and Diggle and Hall [6], to name but a few, where further details and references to particular applications may be found. In the broader context of statistical inverse problems, of which deconvolution is just a special case, are the papers by Koo [15], Fan [8, 10], Mair and Ruymgaart [16], and Cavalier and Tsybakov [5], again to name only a few.

The references mentioned above are representative of a standard framework for the study of statistical deconvolution as well as statistical inverse problems where a well defined theory has emerged. Nevertheless, there does exist a nonstandard yet entirely practical situation to which the standard framework does not apply. This is the case where the independent errors, contaminating the true observations, come from a uniform distribution – the so-called uniform deconvolution problem. This problem is practically of considerable importance because convolution with a uniform distribution corresponds (in signal processing terms) to the blurring that occurs when an optical device undergoes uniform motion with a finite exposure time; see for example Bertero and Boccacci [2]. From the purely statistical point of view, which is the one emphasized here, uniform deconvolution is of special interest due to its unusual features and asymptotics. In particular, while the characteristic function of a uniform distribution declines inversely to it's frequency, it nevertheless oscillates, touching zero at every point on a lattice except at the point corresponding to the zero frequency. Therefore the standard asymptotic results referred to above, do not directly apply to uniform deconvolution. The key issue which we address in this paper is: can one obtain mean integrated squared error bounds in this nonstandard case? We shall show that, with appropriate modifications to the usual estimators, such bounds for uniform deconvolution can be achieved over certain smooth classes of nonperiodic functions. Although other authors have attempted to address this issue with varying degrees of success – see van Es [7], O'Sullivan and Choudhury [17], Hall, Ruymgaart, van Gaans and van Rooij [12], Groeneboom and Jongbloed [11], and Johnstone and Raimondo [14] – for uniform deconvolution in statistics, as well as related statistical inverse problems, our results appear to be the most substantive available thus far for this problem.

We now summarize the contents of this paper. In section 2, we provide the standard theory for deconvolution with smooth (i.e. polynomial) characteristic function of the error and present a standard result for smooth deconvolution of non-periodic functions over Sobolev ellipsoids. This will provide the standard framework. In section 3, we examine the nonstandard problem associated with deconvolution with respect to a uniform error distribution. This problem is nonstandard in the sense that a one sided polynomial bound to the characteristic function is matched with a zero bound on the other side so that standard deconvolution theorems cannot be used to assess the integrated mean-squared error. Nevertheless, we show that it is still possible to derive the same upper bound rate for uniform deconvolution. Motivation for this is discussed in subsection 3.1 while the asymptotic properties of our new procedures are studied in subsection 3.2. The case when the parameter of the uniform distribution is unknown corresponds, for instance, to optical blurring when the exposure time or the speed of uniform motion is not known precisely; this case is dealt with in section 4. Finally, a numerical demonstration is provided in section 6.

We remark that the engineering literature also makes reference to deconvolution under uniform blurring. See, e.g., Bonmassar and Schwartz (1999). When the divisor has zeros, a widely used regularization approach involves bounding the denominator away from zero by adding a positive regularization parameter, the so-called Wiener filtering approach. The procedure we investigate, however, is based on a substantially different approach.

Notation used throughout is as follows. Denote the real line by $\mathbb{R}$ and let $L^2(\mathbb{R})$ be the set of square integrable real valued functions on $\mathbb{R}$. We denote the L^2 norm by $\|f\|^2 = \int |f|^2$ for $f \in L^2(\mathbb{R})$. For $x \in \mathbb{R}$, we denote it's positive part by $x_+ = \max(x,0)$. For order of magnitude we use the Vinogradov notation: for two sequences $\{a_n\}$ and $\{b_n\}$, we symbolize $|a_n| \le c|b_n|$ for some $c > 0$ as $n \to \infty$, by $a_n \ll b_n$ as $n \to \infty$. Also Landau's in probability will be used, so that if the above two sequences are random, then $a_n = O_P(b_n)$ as $n \to \infty$ means $P(|a_n|/|b_n| > c) \to 0$ for some $c > 0$ as $n \to \infty$. If there exist constants $0 < c \le C < \infty$ such that $c|a_n| \le |b_n| \le C|a_n|$ as $n \to \infty$, we shall denote this by $a_n \asymp b_n$ as $n \to \infty$. Expectation will be denoted by $\mathbb{E}$ and variances and covariances by $\mathbb{V}$ar and $\mathbb{C}$ov. Finally, for complex quantities, overlines denote complex conjugation.

2. Smooth deconvolution

In this section we review the known result for deconvolution of non-periodic functions on the real line following the approach of Fan [9].

2.1. Preliminaries

We begin by considering the measurement error problem

$$Y_j = X_j + \varepsilon_j, \tag{2.1}$$

where X_j and ε_j are independent real valued random variables for $j = 1, \dots, n$. Let f_Y, f_X and f_ε denote the densities of Y, X and ε, respectively. Then

$$f_Y(x) = f_X * f_\varepsilon(x), \tag{2.2}$$

where $x \in \mathbb{R}$ and for $f, h \in L^2(\mathbb{R})$, $f * h(x) = \int f(x-y)h(y)dy$ denotes convolution.

The Fourier transform of $f \in L^2(\mathbb{R})$ is defined by

$$\varphi(t) = \int_{\mathbb{R}} f(x)e^{itx}dx, \tag{2.3}$$

where $i^2 = -1$ and $t \in \mathbb{R}$, with Fourier inversion being

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \varphi(t) e^{-itx} dt,$$

where $x \in \mathbb{R}$. The Fourier transform is also referred to as a characteristic function, and the characteristic functions of f_Y, f_X and f_ε will be denoted by φ_Y, φ_X and φ_ε, respectively. Fourier transforming (2.2) gives

$$\varphi_Y(t) = \varphi_X(t)\,\varphi_\varepsilon(t) \tag{2.4}$$

where $t \in \mathbb{R}$.

In the nonparametric context, the parameter of interest is f_X and this must be estimated from the data $Y_1, \ldots, Y_n$. To do this it is assumed that f_ε and hence also φ_ε are known, and we form the empirical characteristic function of $\varphi \equiv \varphi_Y$

$$\varphi_n(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itY_j}$$

for $t \in \mathbb{R}$. Using (2.4) leads to an estimate of φ_X,

$$\hat{\varphi}_X(t) = \frac{\varphi_n(t)}{\varphi_\varepsilon(t)} \tag{2.5}$$

for $t \in \mathbb{R}$, followed by empirical inversion

$$\hat{f}_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\varphi_n(t)}{\varphi_\varepsilon(t)} W(a_n t) e^{-itx} dt \tag{2.6}$$

for $x \in \mathbb{R}$, where $W(t)$ is an appropriate window function, and $a_n \downarrow 0$ an appropriate sequence of tuning (or bandwidth) parameters.

The above is known as the deconvolution problem and (2.6) is known as a deconvolution density estimator. One statistical objective is to try to understand how well $\hat{f}_X$ approximates f_X. Various measures can be used; in this paper we use the L^2 norm and examine how fast $\mathbb{E}\|\hat{f}_X - f_X\|^2$ converges to zero as $n \to \infty$. We note that although other metrics can be used, the L^2 norm is the natural metric when Sobolev ellipsoids are taken as the parameter space, as we do in this paper, since one can then invoke Plancherel's formula.

2.2. Mean integrated squared error bounds

Let $\hat{f}_X$ denote an estimator of f_X and suppose the latter belongs to the Sobolev ellipsoid

$$\Theta_s(Q) = \left\{ f : f \geq 0, \int f = 1, \|f^{(s)}\|^2 \leq Q \right\}, \tag{2.7}$$

where $f^{(s)}$ denotes the $s-$th derivative of f. We say that $\hat{f}_X$ is asymptotically bounded over the class of densities in $\Theta_s(Q)$ if it satisfies

$$\sup_{f_X \in \Theta_s(Q)} \mathbb{E}\left\|\hat{f}_X - f_X\right\|^2 \ll r_n \quad \text{as} \quad n \to \infty, \tag{2.8}$$

where r_n is called the (upper bound) rate of convergence.

The characteristic function, $\varphi_\varepsilon(t)$, $t \in \mathbb{R}$, of the noise or error distribution in a deconvolution problem is called smooth if

$$c_0|t|^{-\gamma} \leq |\varphi_\varepsilon(t)| \leq c_1|t|^{-\gamma}, \text{ as } |t| \to \infty, \tag{2.9}$$

where $0 < c_0 \leq c_1 < \infty$ and $\gamma \geq 0$.

We then have the following which is essentially Theorem 1 of Fan [9], hence we state the result without proof.

Theorem 2.1. *Let $s > 1/2$, $Q > 0$ and suppose $\phi_\varepsilon(t)$ satisfies* (2.9). *Then the estimator* (2.6) *satisfies*

$$\sup_{f_X \in \Theta_s(Q)} \mathbb{E}\left\| \hat{f}_X - f_X \right\|^2 \ll n^{-2s/(2s+2\gamma+1)}$$

as $n \to \infty$.

Remark 2.2. As an example, the Gamma distribution with shape parameter $\alpha > 0$ and scale β has characteristic function given by $(1 - i\beta t)^{-\alpha}$ and so has a tail that declines at the rate $|t|^{-\alpha}$. For this error distribution, Theorem 2.1 says that the fastest achievable rate of convergence for deconvolving the density is $n^{-2s/(2s+2\alpha+1)}$ where s is the number of bounded derivatives which the unknown density is assumed to possess. Note also that for the double-exponential distribution, whose characteristic function is given by $(1+\beta^2t^2)^{-1}$, the upper bound rate is $n^{-s/(2s+5)}$. The Gamma distributions are examples of smooth distributions, in that they satisfy

$$d_0|t|^{-\gamma} \leq |\varphi(t)| \leq d_1|t|^{-\gamma} \tag{2.10}$$

for positive constants d_0, d_1, and $\gamma \geq 0$, as $|t| \to \infty$. Note that the uniform distribution satisfies the right hand side of (2.10) with $\gamma = 1$, but does not satisfy the left hand side.

Remark 2.3. In the proof of the upper bound, only the left inequality of (2.9) is needed. The right inequality is used for the lower bound calculation which in this case can be shown to be the same as the upper bound. The lower bound calculation can be derived using the results of Fan [10].

3. Uniform deconvolution

In this section we pursue a nonstandard, but practically important deconvolution problem. Specifically, in (2.1), we assume the ε_j's are independently uniformly distributed on $[-h,h]$. We shall assume here that $h > 0$ is known so that the convolution (2.2) leads, in obvious notation, to the equation $\varphi(t) \equiv \varphi_Y(t) = \varphi_X(t)\,\varphi_\varepsilon(t)$ among the characteristic functions, where

$$\varphi_\varepsilon(t) = (1/2h)\int_{-h}^{h} e^{itx}\,dx = \frac{\sin(ht)}{ht}\,.$$

For notational simplicity, we select units so that $h = 1$ and $\varphi_\varepsilon(t) = \mathrm{sinc}(t) \equiv \sin(t)/t$, and consequently

$$\varphi_X(t) = \frac{\varphi_Y(t)}{\mathrm{sinc}(t)}\,.$$

The difficulty here, of course, is that $\mathrm{sinc}(t)$ does not satisfy the left side of (2.9) and we therefore cannot apply Theorem 2.1; see also Remark 2.3. In fact it is because

$$0 \le \mathrm{sinc}(t) \le |t|^{-1} \tag{3.1}$$

for all $t \in \mathbb{R}$, with the lower bound attained at integer multiples of π, that this is a nonstandard deconvolution problem. We shall show that despite this nonstandard feature, we can still obtain an estimator that achieves the same rate of convergence as in Theorem 2.1. We first however, provide some motivation.

3.1. Motivation for our estimator

A natural, if naive, proposal for an estimator of the characteristic function of X is

$$\hat{\varphi}_X(t) = \frac{\frac{1}{n}\sum_{j=1}^n e^{itY_j}}{\mathrm{sinc}(t)} \tag{3.2}$$

leading to a proposed Fourier inversion-based density estimate

$$\hat{f}_X(x) = \frac{1}{2\pi n}\int_{-\infty}^{\infty} \frac{\sum_j e^{itY_j}}{\mathrm{sinc}(t)}\, W(a_n t)\, e^{-itx}\, dt\,, \tag{3.3}$$

where $W(t)$ is some appropriate window function, and $a_n \downarrow 0$ an appropriately selected sequence of tuning (or bandwidth) parameters. An immediate problem with the estimator (3.3) is that while the expected value (under the model) of the numerator of the integrand in (3.3) is zero wherever the denominator has a zero, the same does not hold true for the sample values of the numerator. Hence the ratio of the characteristic functions in (3.3) will have 'singularities' at the points $t = \pm\pi, \pm 2\pi, \pm 3\pi, \cdots$ which constitute a lattice except for one point having been removed.

Various devices may be considered in an effort to bridge these singularities. Since characteristic functions are uniformly continuous functions with their roughest point at the origin, are typically smoother away from the origin, and (at least for densities) tend to 0 as $|t| \to \infty$, one might try to establish intervals around the singular points over which the ratio (3.2) of the characteristic functions would not be computed directly, but should instead be estimated by bridging across these intervals by smoothing estimated values of the ratio at locations nearby but outside those intervals. Alternatively, at the zeros of $\mathrm{sinc}(t)$, we could consider applying l'Hôpital's rule to the ratio (3.2); thus at the points $t = k\pi$, where $k = \pm 1, \pm 2, \ldots$, we are led to estimate $\varphi_Y(t)/\{\mathrm{sinc}(t)\}$ as

$$(-1)^k\, \frac{k\pi i}{n} \sum_j Y_j e^{itY_j}\,.$$

These values could then be combined with the bridged estimates mentioned above. We may carry this idea a little further by recalling that l'Hôpital's rule stems from a Taylor expansion argument. This suggests using ratios of Taylor expansions around the singularity points. At the singularities the true values of the constant coefficients of the Taylor expansions are known to be zeros, so it suffices to estimate, say, only the linear, quadratic, and perhaps a few higher coefficients in the numerator. The denominator, on the other hand, is known precisely, and therefore does not need to be expanded. Hence, in suitably small intervals about $t_0 = k\pi$ for $k = \pm 1, \pm 2, \cdots$, the estimator for the ratio of the characteristic functions would have a form such as

$$\frac{1}{n}\,\frac{i(t-t_0)\sum_j Y_j e^{it_0 Y_j} + \frac{1}{2} i^2 (t-t_0)^2 \sum Y_j^2 e^{it_0 Y_j} + \cdots}{\mathrm{sinc}(t)}\,.$$

We may note further that the coefficients of the numerator are correlated so that when correcting one coefficient it should be advantageous to simultaneously (linearly) also correct the others, taking the covariance structure into account. These covariances are not difficult to compute and estimate. We also remark that if Taylor expansion were taken to extremes and the numerator expanded to infinite order at a singularity t_0, except with the first (i.e. the constant) term set equal to zero, we would then just recover the numerator sample characteristic function, but recentered, so that for t in the vicinity of t_0, we would just have replaced the estimated characteristic function $\varphi_n(t)$ by $\varphi_n(t) - \varphi_n(t_0)$. Statistically, this merely incorporates the categorical information in the model that $\varphi(t_0) = \varphi_X(t_0)\,\mathrm{sinc}\,(t_0) = 0$. The final estimator would then just consist of blending together – in the numerator – the $\varphi_n(t) \equiv \frac{1}{n}\sum_{j=1}^n e^{itY_j}$ function with the $\varphi_n(t) - \varphi_n(t_0)$ functions near the singularities.

A different approach might be based on a nonparametric regression-type model

$$\varphi_n(t) = \varphi(t) + e(t) = \varphi_X(t)\varphi_\varepsilon(t) + e(t) \tag{3.4}$$

where the multiplier function $\varphi_\varepsilon(t)$ is known, and where the random function $e(t)$ acts as an error term. For large sample sizes n, this zero mean random function $e(t)$ behaves essentially like a Gaussian process, with a covariance structure that is easily computed and estimated. We could then undertake to fit a smooth function $\varphi_X(t)$ within this model using a procedure that constrains for the smoothness, boundedness (and, to the extent possible, nonnegative definiteness) of $\varphi_X(t)$, taking the covariance structure of the error into account, perhaps using splines or regularized likelihood procedures.

Closer to the approach that we seek to develop here is the idea to adjust, but very smoothly, the estimated $\varphi(t)$ function so that it will have zeros at the $k\pi$ singularity points. In fact, Shannon's Sampling Theorem could be used to produce an especially smooth 'corrector' function, namely

$$\xi(t) = \sum_{k \neq 0} \varphi_n(k\pi)\,\frac{\sin(t-k\pi)}{t-k\pi} \tag{3.5}$$

which would be subtracted from $\varphi_n(t)$. In the sense of being Fourier bandlimited (in the data-domain) (3.5) is the smoothest function which takes on the specified corrected values at the singularities. We remark that even if an infinite number of terms is used in (3.5) the variance of $\xi(t)$ remains bounded, although in practice we might limit the sum to terms corresponding to singularity points actually in sample, i.e. within the segment of the characteristic function used in the Fourier inversion; let these points correspond to $0 \neq |k| \leq K$. Separating out real and imaginary components, and combining terms for the positive and negative k we see that

$$\begin{aligned}
\xi(t) &= \sum_{0\neq|k|\leq K} \varphi_n(k\pi)\,\frac{\sin(t-k\pi)}{t-k\pi} \\
&= \sum_{k=1}^{K} \Re\,\varphi_n(k\pi)\left\{\frac{\sin(t-k\pi)}{t-k\pi}+\frac{\sin(t+k\pi)}{t+k\pi}\right\} \\
&\quad + i\sum_{k=1}^{K} \Im\,\varphi_n(k\pi)\left\{\frac{\sin(t-k\pi)}{t-k\pi}-\frac{\sin(t+k\pi)}{t+k\pi}\right\} \\
&= \sum_{k=1}^{K} \Re\,\varphi_n(k\pi)\sin(t-k\pi))\,\frac{2t}{t^2-k^2\pi^2} \\
&\quad + i\sum_{k=1}^{K} \Im\,\varphi_n(k\pi)\sin(t-k\pi))\,\frac{2k\pi}{t^2-k^2\pi^2}
\end{aligned}$$

so that

$$\xi(t) = 2\sin(t)\sum_{k=1}^{K}(-1)^k\,\frac{t\,\Re\,\varphi_n(k\pi)+k\pi i\,\Im\,\varphi_n(k\pi)}{t^2-k^2\pi^2}\,,$$

where $\Re$ and $\Im$ denote the real and imaginary parts of a complex quantity. Numerical experiments show that this corrector function has exceptional properties in practice and it will turn out that there are good theoretical reasons for this.

In search of an asymptotically optimal corrector function, we shall now try to find one based on the asymptotically normal behaviour of points on the sample characteristic function, and based on the fact that the sample characteristic function is unbiased with a covariance structure that is known and $O_P(n^{-1/2})$ estimable. To begin with, standard computations give

$$n\mathbb{C}\text{ov}\,(\varphi_n(s),\varphi_n(t)) = \varphi(s-t)-\varphi(s)\overline{\varphi}(t)$$

and, using $\Re\,\varphi_n(t) = [\varphi_n(t)+\varphi_n(-t)]/2$ and $\Im\,\varphi_n(t) = [\varphi_n(t)-\varphi_n(-t)]/2i$, we obtain

$$n\mathbb{C}\text{ov}\,(\Re\,\varphi_n(s),\Re\,\varphi_n(t)) = \frac{1}{2}\,[\Re\,\varphi(s-t)+\Re\,\varphi(s+t)]-\Re\,\varphi(s)\,\Re\,\varphi(t) \tag{3.6}$$

$$n\mathbb{C}\text{ov}\,(\Re\,\varphi_n(s),\Im\,\varphi_n(t)) = \frac{1}{2}\,[\Im\,\varphi(s-t)+\Im\,\varphi(s+t)]-\Re\,\varphi(s)\,\Im\,\varphi(t) \tag{3.7}$$

$$n\mathbb{C}\text{ov}\,(\Im\,\varphi_n(s),\Im\,\varphi_n(t)) = \frac{1}{2}\,[\Re\,\varphi(s-t)-\Re\,\varphi(s+t)]-\Im\,\varphi(s)\,\Im\,\varphi(t)\,. \tag{3.8}$$

Observe now that for s and t different, but both of form $k\pi$ for integers $k \neq 0$, many of the terms on the right in (3.6) - (3.8) vanish; in fact, the real and imaginary parts of $\varphi_n(k\pi)$ for $k = 1, 2, \cdots$, are all uncorrelated with common variance $1/2$. Now suppose that we seek, in a simple linear way, to correct the sample value of $\varphi_n(t)$ for some t on the basis of information contained in the observed values of φ_n at the known true zeros $t = k\pi$, $k \neq 0$. Because φ_n is an average of bounded iid random functions, it is asymptotically normal at finite collections of points so it is natural to consider regression-based correction. Since the predictors here are uncorrelated, the corrections can be determined for each predictor individually, and then just added together. Now in the standard regression problem $Y = \beta X + e$ for zero mean random variables X, Y and e, we would select $\beta = \mathbb{C}\mathrm{ov}(X,Y)/\mathbb{V}\mathrm{ar}(X)$, and the correction would be taken to be βX. Hence the correction to $\Re\,\varphi_n(t)$ due to the observed departure of $\Re\,\varphi_n(k\pi)$ from zero would be given by

$$\{\Re\,\varphi(t-k\pi)+\Re\,\varphi(t+k\pi)\}\cdot\Re\,\varphi_n(k\pi)\,.$$

Likewise, we obtain the correction to $\Re\,\varphi_n(t)$ due to the observed non-zero value of $\Im\,\varphi_n(k\pi)$, and the corrections to $\Im\,\varphi_n(t)$ due to the observed real and imaginary parts of $\varphi_n(k\pi)$. Putting together these four terms, the overall correction to $\varphi_n(t)$ due to the nonzero value of $\varphi_n(k\pi)$ is found to be

$$\begin{aligned}&\big[\{\Re\,\varphi(t-k\pi)+\Re\,\varphi(t+k\pi)\}\Re\,\varphi_n(k\pi)+\{\Im\,\varphi(t-k\pi)+\Im\,\varphi(t+k\pi)\}\,\Im\,\varphi_n(k\pi)\big]\\&\quad+i\,\big[\{-\Im\,\varphi(t-k\pi)+\Im\,\varphi(t+k\pi)\}\Re\,\varphi_n(k\pi)+\{\Re\,\varphi(t-k\pi)-\Re\,\varphi(t+k\pi)\}\Im\,\varphi_n(k\pi)\big]\end{aligned}$$

which simplifies to

$$\varphi_n(k\pi)\cdot\varphi(t-k\pi)+\overline{\varphi_n(k\pi)}\cdot\overline{\varphi(t+k\pi)}\,.$$

It follows that the regression-based corrector function is given by

$$\xi(t)=\sum_{k\neq 0}\varphi_n(k\pi)\cdot\varphi(t-k\pi)\,. \tag{3.9}$$

Since $\varphi=\varphi_X\times\mathrm{sinc}$ and sinc $\downarrow 0$, and also since $\varphi_X\downarrow 0$, typically at good rates, we see that for each k the correction for $\varphi_n(k\pi)$ affects $\varphi_n(t)$ only for t in the vicinity of $k\pi$. Therefore in typical applications, only a few $k-$terms need to be preserved in the sum (3.9). Of course, the function $\varphi\equiv\varphi_Y$ in (3.9) is unknown, but it can, in the first instance, be estimated empirically from the Y_j's, and in the second instance, it can be estimated by multiplying an initial estimate of φ_X by the sinc function, a procedure which can be iterated. Remarkably, however, it will turn out that the fact that φ in the corrector function is unknown and has to be estimated does not affect the asymptotic properties of the resulting estimator. Indeed, we shall see that even if the φ_X component of $\varphi=\varphi_X\times\mathrm{sinc}$ in the corrector function is not estimated consistently, the convergence rate of the resulting estimator is unaffected.

It turns out that this regression-based correction procedure is highly effective in practice. Furthermore, the similarity in character of (3.9) and (3.5) in part explains the high numerical effectiveness that is observed for the Shannon corrector; this occurs largely because φ is essentially just the sinc function for values of t around the origin.

We can now write down our final proposal for an estimator of f_X, namely

$$\hat{f}_X(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\Phi_n(t)}{\mathrm{sinc}(t)}\,W(a_nt)\,e^{-itx}\,dt\,, \tag{3.10}$$

where

$$\Phi_n(t)=\varphi_n(t)-\sum_{k\neq 0}\varphi_n(k\pi)\cdot\varphi^{\dagger}(t-k\pi), \tag{3.11}$$

is a modified empirical characteristic function. The $\varphi^\dagger$ here is some characteristic function satisfying $\varphi^\dagger(t) = \varphi_X^\dagger(t)\,\mathrm{sinc}(t)$, where $\varphi_X^\dagger$ is itself some characteristic function. It will turn out that for rate bounds $\varphi_X^\dagger$ does not need to be selected to consistently estimate φ_X; for example we may select $\varphi_X^\dagger$ to be Gaussian with variance matching the sample variance of the data, less the known variance for the uniform error term.

3.2. Mean integrated squared error bound

Our main result for uniform deconvolution can now be stated as

Theorem 3.1. *Suppose φ_ε is uniform, and that X and Y have finite variance. Then the estimator* (3.10) *satisfies*

$$\mathbb{E}\,\big\|\hat{f}_X - f_X\big\|^2 \ll n^{-2s/(2s+3)},$$

for $s > 1/2$ as $n \to \infty$, provided only that $\varphi^\dagger(t) = \varphi_X^\dagger(t)\,\mathrm{sinc}(t)$ with $\varphi_X^\dagger(t)$ being a characteristic function possessing a variance.

3.3. Proof of Theorem 3.1

To determine the integrated mean squared error of our estimator, we first apply Plancherel's formula to obtain:

$$2\pi\,\mathbb{E}\int_{-\infty}^{\infty}\big|\hat{f}_X(x) - f_X(x)\big|^2\,dx = \int_{-\infty}^{\infty}\mathbb{E}\,\left|\frac{\Phi_n(t)}{\mathrm{sinc}(t)}\,W(a_n t) - \varphi_X(t)\right|^2 dt\,, \tag{3.12}$$

where

$$\Phi_n(t) = \varphi_n(t) - \sum_{k\neq 0}\varphi_n(k\pi)\cdot\varphi^\dagger(t-k\pi) \tag{3.13}$$

and

$$\varphi_n(t) = \frac{1}{n}\sum_j e^{itY_j}\,.$$

The function $\Phi_n(t)$ is just the numerator in (3.10) written using $\varphi^\dagger(t) \equiv \varphi_X^\dagger(t)\,\mathrm{sinc}(t)$ in place of $\varphi(t) \equiv \varphi_X(t)\,\mathrm{sinc}(t)$ in the corrector term (3.9). We do this firstly because $\varphi(t)$ appearing in the corrector function (3.9) is not actually known and so needs to be estimated, and also because we will seek to understand the consequences of estimating incorrectly the φ function in the corrector. For the moment we treat the $\varphi^\dagger(t)$ terms as being nonstochastic.

Next, we decompose (3.12) as a sum of two terms, namely the integrated squared bias

$$2\pi\int\big|\mathbb{E}\hat{f}(x) - f(x)\big|^2\,dx = \int\left|\frac{\mathbb{E}\Phi_n(t)}{\mathrm{sinc}(t)}W(a_n t) - \varphi_X(t)\right|^2 dt \tag{3.14}$$

and the integrated variance

$$2\pi \int \mathbb{E}\left|\hat{f}(x) - \mathbb{E}\hat{f}(x)\right|^2 dx = \int \mathbb{E}\left|\frac{\Phi_n(t)}{\operatorname{sinc}(t)} W(a_n t) - \varphi_X(t) W(a_n t)\right|^2 dt\,. \tag{3.15}$$

The bias term (3.14) may be written as

$$\int |\varphi_X(t)|^2 \, |W(a_n t) - 1|^2 \, dt\,,$$

and if W is just the boxcar function $W(t) = 1$ on $[-1, 1]$ and 0 elsewhere, it becomes

$$\int_{|t|>a_n^{-1}} |\varphi_X(t)|^2 \, dt\,.$$

To compute the variance term (3.15) we shall first require to evaluate

$$\begin{aligned}
&n\mathbb{V}\mathrm{ar}\,(\Phi_n(t)) \\
&\quad = \; n\mathbb{C}\mathrm{ov}\left(\varphi_n(t) - \sum_{k\neq 0} \varphi_n(k\pi)\varphi^\dagger(t-k\pi),\; \varphi_n(t) - \sum_{\ell\neq 0} \varphi_n(\ell\pi)\varphi^\dagger(t-\ell\pi)\right) \\
&\quad = \; n\mathbb{V}\mathrm{ar}\,(\varphi_n(t)) - 2\Re \sum_{k\neq 0} \mathbb{C}\mathrm{ov}\,(\varphi_n(t), \varphi_n(k\pi))\,\overline{\varphi^\dagger(t-k\pi)} \\
&\qquad + n \sum_{k\neq 0}\sum_{\ell\neq 0} \varphi^\dagger(t-k\pi)\mathbb{C}\mathrm{ov}\,(\varphi_n(k\pi), \varphi_n(\ell\pi))\,\overline{\varphi^\dagger(t-\ell\pi)} \\
&\quad = \; \left\{1 - |\varphi(t)|^2\right\} - 2\Re \sum_{k\neq 0} \varphi(t-k\pi)\overline{\varphi^\dagger(t-k\pi)} + \sum_{k\neq 0} \left|\varphi^\dagger(t-k\pi)\right|^2 .
\end{aligned}$$

This may be rearranged as

$$\begin{aligned}
&n\mathbb{V}\mathrm{ar}\,(\Phi_n(t)) \\
&\quad = \left\{1 - \sum_{k=-\infty}^{\infty} |\varphi(t-k\pi)|^2\right\} \\
&\qquad + \sum_{k=-\infty}^{\infty} \left|\varphi(t-k\pi) - \varphi^\dagger(t-k\pi)\right|^2 - \left|\varphi(t) - \varphi^\dagger(t)\right|^2 .
\end{aligned} \tag{3.16}$$

We shall need to understand the behaviour of $n\mathbb{V}\mathrm{ar}\,(\Phi_n(t))\,/\operatorname{sinc}^2(t)$ and now undertake to show that this function is bounded by a polynomial of degree 2:

Lemma 3.2. *Suppose* $\mathbb{V}\mathrm{ar}X, \mathbb{V}\mathrm{ar}Y < \infty$. *Then*

$$n\mathbb{V}\mathrm{ar}\left(|\Phi_n(t))\,/\operatorname{sinc}^2(t)\right| \le c_0 + c_1|t| + c_2|t|^2$$

for some $c_0, c_1 \in \mathbb{R}$, *and* $c_2 > 0$ *for all* $t \in \mathbb{R}$.

Proof. To prove this, first note that each of the three terms in (3.16) is continuous, and when divided by $\operatorname{sinc}^2(t)$, each of the three resulting terms respectively tend to 0 as $t \to 0$.

Furthermore, the third of these terms, i.e.

$$\left|\varphi(t)-\varphi^{\dagger}(t)\right|^2\left(\frac{t}{\sin(t)}\right)^2=\left|\varphi_X(t)-\varphi_X^{\dagger}(t)\right|^2$$

and is therefore bounded by 2. We next observe that the other two of the three terms in (3.16) are each π-periodic. Since the $\sin^2(t)$ occurring in $\operatorname{sinc}^2(t)=(\sin(t)/t)^2$ is also π-periodic the first two terms of $n\mathbb{V}\mathrm{ar}\left(\Phi_n(t)\right)/\operatorname{sinc}^2(t)$ will then just be π-periodic functions multiplied by t^2. Therefore, as far as these two terms are concerned, we need only to establish that their limits are finite also at $t=\pi$ and it then will follow that the function $n\mathbb{V}\mathrm{ar}\left(\Phi_n(t)\right)/\operatorname{sinc}^2(t)$ is bounded by a polynomial of degree 2.

To examine the first component then, we first isolate from it the term

$$\left(1-|\varphi(t-\pi)|^2\right)\left(\frac{t}{\sin(t)}\right)^2. \tag{3.17}$$

Here we will need to assume that X and hence also Y possess variances. Then since $|\varphi(u)|^2$ is the characteristic function of the symmetrized version of Y it will be asymptotic to $1-\frac{1}{2}(2\sigma_Y^2)u^2$ as $u\to 0$ where σ_Y^2 is the variance of Y. Hence as $t\to\pi$ (3.17) will be asymptotic to $\sigma_Y^2(t-\pi)^2(t/\sin(t))^2$ and consequently will approach the constant $\pi^2\sigma_Y^2$. The remaining terms of the first component, namely

$$\sum_{k\neq 1}|\varphi(t-k\pi)|^2\left(\frac{t}{\sin(t)}\right)^2 = \sum_{k\neq 1}|\varphi_X(t-k\pi)|^2\left(\frac{\sin(t-k\pi)}{t-k\pi}\right)^2\left(\frac{t}{\sin(t)}\right)^2;$$

as $t\longrightarrow\pi$ this approaches

$$\sum_{k\neq 1}|\varphi_X((1-k)\pi)|^2\left(\frac{\pi}{\pi-k\pi}\right)^2\leq\sum_{k\neq 1}\left(\frac{1}{1-k}\right)^2=\frac{\pi}{3}$$

and therefore is bounded.

Likewise we first isolate the term

$$\left|\varphi(t-\pi)-\varphi^{\dagger}(t-\pi)\right|^2\left(\frac{t}{\sin(t)}\right)^2$$

appearing in the second component. This term equals

$$\left|\varphi_X(t-\pi)-\varphi_X^{\dagger}(t-\pi)\right|^2\left(\frac{\sin(t-\pi)}{t-\pi}\right)^2\left(\frac{t}{\sin(t)}\right)^2. \tag{3.18}$$

But near $u=0$ we have $\varphi_X(u)-\varphi_X^{\dagger}(u)$ asymptotic to $i(\mu_X-\mu_X^{\dagger})u$, where μ_X and $\mu_X^{\dagger}$ are the means corresponding to φ_X and $\varphi_X^{\dagger}$ so that as $t\to\pi$ (3.18) approaches $\pi(\mu_X-\mu_X^{\dagger})^2$. The remaining terms of the second component sum to

$$\sum_{k\neq 1}\left|\varphi_X(t-k\pi)-\varphi_X^{\dagger}(t-k\pi)\right|^2\left(\frac{\sin(t-k\pi)}{t-k\pi}\right)^2\left(\frac{t}{\sin(t)}\right)^2,$$

and as $t \to \pi$ this approaches

$$\sum_{k\neq 1}\left|\varphi_X((1-k)\pi)-\varphi_X^{\dagger}((1-k)\pi)\right|^2\left(\frac{1}{1-k}\right)^2 \leq 2\sum_{k\neq 1}\left(\frac{1}{1-k}\right)^2=\frac{2\pi}{3}$$

which again is bounded. Hence altogether we have established that the term in question is bounded by a polynomial of degree 2. □

Returning now to our proof of the Theorem, by Lemma 3.2, we have that

$$\mathbb{E}\left\|\hat{f}_X-\mathbb{E}\hat{f}_X\right\|^2 \leq \frac{\text{const}}{n}\int_{-a_n^{-1}}^{a_n^{-1}} t^2 dt \ll \frac{1}{na_n^3} \quad .$$

On the other hand

$$\begin{aligned}
\left\|\mathbb{E}\hat{f}_X-f_X\right\|^2 &= \int_{[-a_n^{-1},a_n^{-1}]^c}|\phi_X(t)|^2 dt \\
&= \int_{[-a_n^{-1},a_n^{-1}]^c} t^{-2s}t^{2s}|\phi_X(t)|^2 dt \\
&\leq a_n^{2s}\int_{[-a_n^{-1},a_n^{-1}]^c} t^{2s}|\phi_X(t)|^2 dt \\
&\leq a_n^{2s}\int_{-\infty}^{\infty} t^{2s}|\phi_X(t)|^2 dt \\
&\ll a_n^{2s}
\end{aligned}$$

since we are assuming $f_X \in \Theta_s(Q)$. Putting the above two together, we have

$$\mathbb{E}\left\|\hat{f}_X-f_X\right\|^2 \ll \frac{1}{na_n^3}+a_n^{2s}$$

as $n\to\infty$. Choosing $a_n \asymp n^{-1/(2s+3)}$ gives the lower bound for the right hand side of the above, and hence

$$\mathbb{E}\left\|\hat{f}_X-f_X\right\|^2 \ll n^{-2s/(2s+2+1)}$$

as $n\to\infty$.

Remark 3.3. Three useful observations are in order. Firstly, upon combining the last two terms in (3.16) into a single sum it emerges that the choice $\varphi^{\dagger}=\varphi$ in (3.13) minimizes the variance and hence the mean square error of our estimator. (Note that $\varphi^{\dagger}$ does not appear in the bias component.) This may be considered as an alternative to the derivation of the optimal regression-based corrector carried out in section 3.1; alternately, it may be viewed as a confirmation of that result. Secondly, it is also clear from the derivation here that asymptotic rate bound does not require the optimal selection $\varphi^{\dagger}=\varphi$; in fact it suffices that $\varphi^{\dagger}(t)$ equal the $\text{sinc}(t)$ function multiplied by any characteristic function $\varphi_X^{\dagger}(t)$ which possesses a finite variance; we shall always impose this requirement on $\varphi^{\dagger}(t)$. Thirdly, this observation explains the exceptionally good behaviour which is observed in numerical experiments using the Shannon corrector (3.5); it corresponds to taking $\varphi^{\dagger}(t)$ equal to $\text{sinc}(t)$ multiplied by the degenerate characteristic function which identically equals 1. As a practical point, we mention that the corrector function needs primarily to contain just those terms where k corresponds to singularities within the interval over which the integral (3.10) is evaluated – i.e., where the weight function W is nonzero – and may therefore be taken to have only a finite number of terms. Nevertheless the sinc function does decline rapidly enough that the infinite sum also converges. Finally, we remark that our arguments here have required, in a rather fundamental way, that X have finite variance.

Remark 3.4. While any finite variance choice for $\varphi_X^\dagger$ leads to rate bounds for our estimator, the choice $\varphi_X^\dagger = \varphi_X$ also provides asymptotically the best constant for that rate for our estimator. In practice, when optimality of that constant is desirable one would normally replace $\varphi_X^\dagger$ by an estimate of φ_X.

4. Estimating the uniform parameter

In this section we show that when the parameter h of the $U[-h,h]$ uniform distribution for the error terms is unknown it may, under broad conditions, be estimated consistently with an $O_P(1/\sqrt{n})$ rate. Because this is substantially faster than the convergence rate of the deconvolution density estimator the case of unknown h does not materially affect the asymptotic behaviour of the estimators.

Observe firstly that in order for h to be identifiable from the characteristic function φ of Y it is necessary that the φ_X component should not admit any uniform distribution factors (in the convolution algebra for distributions) other perhaps than additional factors of $U[-h,h]$. We must also require that the location of any additional zeros of φ_X not interfere with our ability to asymptotically identify the lattice $k\pi/h$, $k = \pm 1, \pm 2, \cdots$ formed by the zeros of φ_U; this condition is, of course, very mild. Finally, in order that we be able to estimate h with rate $O_P(1/\sqrt{n})$ from an estimate of the true zero of φ occurring at π/h we shall require the behaviour of φ_X to be such that φ is differentiable at π/h with a nonzero derivative there. It is not a material restriction to assume that the first zero of φ in fact corresponds to the first zero of φ_U, especially if the data has been centered prior to analysis, and purely for the sake of brevity we shall assume this to be the case. (The characteristic function for noncentered data will have an $e^{i\mu t}$ type of factor which produces zeros.) In this case the derivative of φ will in fact be negative at π/h. Note also that an empirical characteristic function will typically not have exact zeros on its real and imaginary parts simultaneously; we therefore choose to work primarily with zeros of the real part. Indeed for nearly symmetrical distributions the imaginary part of the empirical characteristic function will typically contain less information about h than the real part.

The parameter h can be estimated in may ways, and the most efficient way to do so will necessarily depend on the context. For example, using a small value of J, we may consider using a test statistic such as

$$\sum_{j=1}^{J} |\varphi_n(j\pi/h)|^2 \tag{4.1}$$

which will have the distribution of a $\chi^2_{2J}/2$ variable when the value of h is correct; this observation leads to a range of plausible values for h. (When h is correct, the real and imaginary terms in (4.1) are asymptotically independent normals with variances 1/2.) Minimizing this statistic over the 'inside' range for h will lead to a consistent estimate of h. For reasons of power/efficiency, it may be preferable to use only the real components of the empirical characteristic function points in (4.1).

In fact, often the first zero, $\hat{t}_0$ say, of the real part of $\varphi_n(t)$ is all that is needed, and h can then be estimated from it as $\hat{h} = \pi/\hat{t}_0$. The estimation of such first zeros has been considered in Welsh [19], Heathcote and Hüsler [13], and by Braker and Hüsler [3]. In particular, [19] considered the numerical determination of $\hat{t}_0$ and proved its almost sure convergence to the true first zero, t_0. Under mild conditions, Theorem 3.1 of [13], shows that $\hat{t}_0$ is asymptotically normal with mean t_0 and variance $c/\sqrt{n}$, where the constant c is given by $c = \sigma(t_0)/|u'(t_0)|$ where $\sigma^2(t) = \frac{1}{2}[1 + u(2t) - 2u^2(t)]$, $u'(t)$ is the derivative of $u(t)$, and $u(t) = \Re\varphi(t)$. Note that in our context $\sigma(t_0) = 1/\sqrt{2}$ and $u'(t_0) = (h/\pi)u_X(t_0)$ where $u_X(t)$ is just the real part of φ_X.

The work of [13] misstates a moment condition, and is based on a rather complicated weak convergence argument. The correct result may be derived in a much simpler fashion as follows. Let

$$U_n(t) = \frac{1}{n}\sum_{j=1}^{n}\cos(X_j t)$$

be the real part of the empirical characteristic function and

$$u(t) = \mathbb{E}\cos(Xt)$$

be the real part of the characteristic function. Let

$$T_n = \inf\{t : U_n(t) = 0\}$$

be the first zero of the empirical characteristic function and

$$t_0 = \inf\{t : u(t) = 0\}$$

be the first zero of the characteristic function. When t_0 is properly behaved (as it is in our application), Welsh [19] showed that $T_n \to t_0$ almost surely. Now consider, for each n, the Taylor expansion

$$U_n(\hat{t}_n) = U_n(t_0) + (\hat{t}_n - t_0)U_n'(t_n^*) \tag{4.2}$$

where t_n^* lies between t_0 and $\hat{t}_n$. If, for each n, we select $\hat{t}_n$ to be the root T_n, then the left sides of each of the equations (4.2) will be zero and so can be solved to give

$$T_n = t_0 - \frac{U_n(t_0)}{U_n'(t_n^*)} \tag{4.3}$$

where now $\hat{t}_n$ lies between t_0 and T_n and hence tends to t_0 by [19]'s result.

Now for a fixed t,

$$U_n'(t) = -\frac{1}{n}\sum_{j=1}^{n}X_j\sin(X_j t)$$

and this converges to

$$u'(t) = -\mathbb{E}\{X\sin(Xt)\}\,.$$

But since $X\sin(Xt)$ is uniformly bounded by $|X|$, then provided $\mathbb{E}|X| < \infty$ we will have that $u'(t)$ is continuous in t using the Dominated Convergence Theorem. Furthermore a Uniform Law of Large Numbers will apply to give that $U_n'(t)$ tends to $u'(t)$ uniformly over any finite interval in t; see, e.g., Andrews [1]. Consequently the denominator in (4.3) will converge to $u'(t_0)$ in view of Welsh's result. We therefore find that T_n is asymptotically normal:

$$\sqrt{n}\,(T_n - t_0) \to^d N\left(0,\, \frac{1}{2\,|u'(t_0)|^2}\right)$$

where $\to^d$ is convergence in distribution. Note that our argument relies on the finiteness of $\mathbb{E}|X|$ in a fundamental way; this moment condition cannot be relaxed but no others are required. Since h is $O_P\,(1/\sqrt{n})$ estimable, using such an estimated value for h in the deconvolution algorithm does not affect its asymptotic properties.

5. Non-homogeneous case

The nonhomogeneous case in which our observations have been convolved with $U[-h_j, h_j]$ uniform errors using a different (but known) h_j in each case can be handled, for example, as follows. Each $e^{itY_j}/\mathrm{sinc}(h_j t)$ is an unbiased estimator for φ_X and these estimators can be judiciously combined as

$$\hat{f}_X(x) = \frac{1}{2\pi n}\sum_{j=1}^{n}\int_{-\infty}^{\infty} \Phi_{n,j}(t) W(a_n t)\, e^{-itx}\, dt\,, \tag{5.1}$$

where

$$\Phi_{n,j}(t) = \frac{e^{itY_j} - \sum_{k\neq 0} e^{i(k\pi/h_j)Y_j}\,\varphi_X^{\dagger}(t - k\pi/h_j)\mathrm{sinc}(t - k\pi/h_j)}{\mathrm{sinc}(t - k\pi/h_j)}$$

and where the function $\varphi_X^{\dagger}$ can be estimated using methods analogous to before. While this appears to be a reasonable estimator in this context, we note that the terms in it are not identically distributed and this increases the complexity of any asymptotic considerations.

6. Numerical experiment

The uniform deconvolution procedures developed in section 3 were implemented using the S-Plus statistical software package and extensive experimentation showed our methods to be numerically well-behaved. As an illustration of the basic workings of our methods, parts (a) – (h) of Figure 1 are based on a simulated sample from a normal mixture distribution in which 50% of the observations come from a normal with mean 0.2 and standard deviation 0.2 and 50% of the observations come from a normal with mean -0.3 and standard deviation 0.1. These 'X' observations were then convolved with independent errors from a uniform

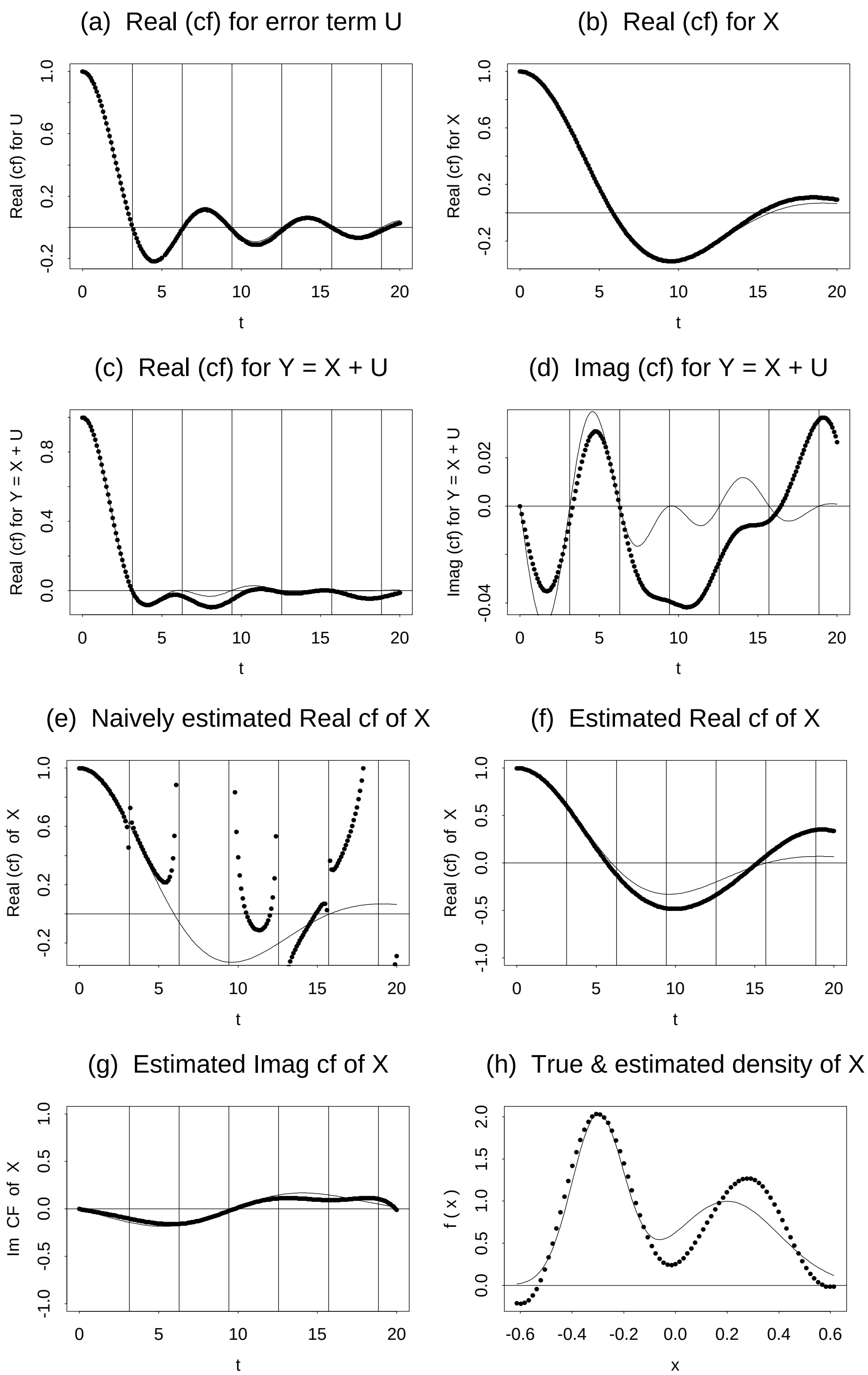

Figure 1. An experiment in uniform deconvolution based on a sample size of $n = 1000$ and a signal from a normal mixture model.

distribution on $[-1,1]$. The parameter values for f_X were chosen for clarity of the graphical presentation. Because we are dealing with a fully nonparametric problem, and so as to ensure that the key features of our methods are fully evident, we used a large random sample of size $n = 1000$ in producing Figure 1. For purposes of this illustration, we also eliminated one minor source of variation by taking exactly 500 of the 1000 observations from each of the two respective components of the normal mixture.

Part (a) of Figure 1 shows the real parts of the true and estimated characteristic functions for the uniform error terms. Here, as elsewhere in Figure 1, estimates are shown as dots, while true values are shown as a solid line. Here these two curves are rather close together and therefore cannot be distinguished. For convenience, we have superimposed (both here and in other parts of this figure) horizontal lines at $t = \pi, 2\pi, \cdots$ which are the singularity points of the uniform $U[-1,1]$ error distribution. Part (b) of the figure shows the real parts of the true and estimated characteristic functions for the 'signal' variable X. Here again, the true and estimated curves virtually coincide except at the larger values of the argument t. This figure is provided only for illustrative purposes; in practice we would, of course, not have direct access to values of the X variables.

Part (c) of Figure 1 shows the real parts of the true and estimated characteristic functions for the Y's, i.e., for the convolved data. The two curves are again quite close here except for small differences that occur at values of t between about 5 and 10. Part (d) is the same as part (c) except based on the imaginary parts of the characteristic functions. The numerical values for the imaginary parts may be noted to be fairly small; this is owing to the fact that the density for X is not unduly far from being symmetrical about the origin.

Part (e) of Figure 1 shows what happens if one tries to naively estimate the characteristic function of X by simply dividing the estimated characteristic function of Y by the known characteristic function for the uniform errors. This plot is for the real components of the characteristic function. As we approach each of the 'singularity' points $t = \pi, 2\pi, \cdots$ then, depending upon whether the real part of the sample characteristic function differs from 0 by being positive or negative there, the estimated characteristic function of X will approach $+\infty$ from one side and $-\infty$ from the other side of the singularity. The grid-spacing for t used here only allows this effect to be seen clearly from the left of second singularity, as well as from the right and left, respectively, of the third and fourth singularities. The values of the estimates between the second and third singularities here all fall below the visible part of the plot.

A number of methods for uniform deconvolution have been discussed in this paper. The method illustrated in parts (f) and (g) for estimating the real and imaginary components of the characteristic function of X is based on simply using a corrector function obtained from multiplying the sinc function corresponding to the known characteristic function of the uniform error with the characteristic function of a normal distribution whose variance is the sample variance of the Y data less the known variance of the uniform error distribution. In other words, a normal distribution was used here in place of the $\varphi_X^\dagger$ function, and its variance was taken to be an estimate of the variance of X – a particularly simple procedure to implement. It is seen that in this instance the real characteristic function is underestimated slightly for t between 5 and 15 and overestimated somewhat for t above 15; the imaginary characteristic function is estimated fairly accurately here.

Finally, part (h) of Figure 1 gives the Fourier inversion of the estimated characteristic function of X. Tukey tapering, based on cosine half-bells, was applied prior to invoking an FFT-based algorithm. In this instance it is seen that the two bumps of the true density function (solid line) are estimated rather faithfully; the left part of the density is in fact estimated quite accurately here although the estimated density does fall below 0 at the left; the right part of the density here is less well estimated, but is still accurate in its broad features. Of course, the density estimate for the convolved Y data (not shown here) does not evidence the two distinct bumps that occur in the density of X.

Acknowledgements

Andrey Feuerverger and Peter Kim were supported by the Natural Sciences and Engineering Research Council (Canada) grants OGP9133 and OGP46204, respectively.

References

Andrews, D., 1992. Generic uniform convergence. *Econometric Theory*, 8, 241–247.

Bertero, M., Boccacci, P., 1980. *Introduction to Inverse Problems in Imaging*. Institute of Physics, Bristol.

Braker, H.U., Hüsler, J., 1991. On the first zero of an empirical characteristic function. *J. Appl. Probab.*, 28, 593–601.

Carroll, R.J., Hall, P., 1988. Optimal rates of convergence for deconvoluting a density. *J. Amer. Statist. Assoc.*, 83, 1184–1186.

Cavalier, L., Tsybakov, A., 2002. Sharp adaptation for inverse problems with random noise. *Probab. Theory Relat. Fields*, 123, 323–354.

Diggle, P.J., Hall, P., 1993. A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc.*, Ser. B 55, 523–531.

van Es, A.J., 1991. Uniform deconvolution: nonparametric maximum likelihood and inverse estimation. *Nonparametric functional estimation and related topics (Spetses, 1990)*, 191–198. NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 335, Kluwer Acad. Publ., Dordrecht.

Fan, J., 1991. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19, 1257–1272.

Fan, J., 1992. Global behaviour of deconvolution kernel estimates. *Statist. Sinica*, 1, 541–551.

Fan, J., 1993. Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.*, 21, 600–610.

Groeneboom, P., Jongbloed, G., 2003. Density estimation in the uniform deconvolution model. *Statist. Neerleandica*, 57, 136–157.

Hall, P., Ruymgaart, F., van Gaans, O., van Rooij, A., 2001. Inverting noisy integral equations using wavelet espansions: A class of irregular convolutions. In *State of the Art in Probability and Statistics: Festschrift for Willem R. van Zwet* 533–546, IMS, Beachwood, OH.

Heathcote, C.R., Hüsler, J., 1990. The first zero of an empirical characteristic function. *Stochastic Process. Appl.*, 35, 347–360.

Johnstone, I.M., Raimondo, M., 2004. Periodic boxcar deconvolution and diophantine approximation. *Ann. Statist.*, 32, 1781–1804.

Koo, J.-Y., 1993. Optimal rates of convergence for nonparametric statistical inverse problems. *Ann. Statist.*, 21, 590–599.

Mair, B.A., Ruymgaart, F.H., 1996. Statistical inverse estimation in Hilbert scales. *SIAM J Appl Math*, 56, 1424–1444.

O'Sullivan, F., Choudhury, K., 2001. An analysis of the role of positivity and mixture model constraints in Poisson deconvolution problems. *J. Comput. Graph. Statist.*, 10, 673–696.

Stefanski, L., Carroll, R.J., 1990. Deconvoluting kernel density estimators. *Statistics*, 21, 169–184.

Welsh, A.H., 1986. Implementing empirical characteristic function procedures. *Statist. Probab. Lett.*, 4, 65–67.

Zhang, C.H., 1990. Fourier methods for estimating mixing densities and distributions. *Ann. Statist.*, 18, 806–830.